Chapter 2

Paper II

Psychological Measurement of Individual Differences

1) Characteristics of a good Psychological Test (15)/(10)
2) Construction of a Psychological Test (20)
3) Reliability of the test (15)
4) Validity of the test (15)
5) Relationship between Reliability & Validity (10)
6) Limitations of Psychological Tests (15)
7) Uses and misuses of Psychological Tests (10)
8) Ethical issues in Psychological Testing (15)
9) Ethical issues in Research (15)
10) Types of Psychological Testing (10)

*Psychological Test

Instrument to measure ind. differences using standard procedures with as much precision & accuracy as possible.

* Anne Anastasi

Psychological Test is a standardised & an objective measure of a sample of behaviour

.....

Set of items designed to measure characteristics of human beings that pertain to behaviour.

*sample of behaviours

(i) for particular people in the population
(ii) for particular behaviour that make a trait

* Cronbach

A psychological test is a systematic procedure for observing person's behaviour & describing it with the help of numerical scales or fixed categories.

Attempt is to quantify the behaviour for better prediction & control.

Scanned by CamScanner

SHOUT
SHARE
IT

Characteristics of a good Psychological Test

(i) Reliability (consistency)
(ii) Validity (truthfulness & purity)
(iii) Objectivity
(iv) Norms
(v) Practicality

(i) Reliability is the consistency of the test reflected in the reproducibility of the scores.

Reliability of the test indicates the degree to which the test is free from internal defects of standardisation.

(ii) Validity
- ↳ Purity of the test
- ↳ degree to which the test measures what it intends to measure and what it claims to measure
- ↳ whether a test does what it was supposed to do

* Reliability is the self-correlation of the test whereas validity of the test is the correlation of the test scores with an external independent criteria.

Self-correlation (within items of a test)

(iii) Objectivity

Degree to which the test is free from subjectivity

(i) Items of the test (questions ~~defined~~ used in the test)
→ Item objectivity

* Objectivity depends on the objectivity of the items used in the test.

(ii) Objectivity of the scoring system

* Item Objectivity

(i) Items should be phrased in such a manner that they present no ambiguity in their meaning to those who write the test as well as to those who score it.

Scanned by CamScanner

(ii) Items must have moderate degree of difficulty

(very difficult question → intimidate the test taking person → guesses start

(very easy → no interest in the test)

* Objectivity of the scoring system

(i) It should be well defined so that consistency can be obtained (i) when the same test is scored by the same scorer at diff. times.

(ii) When the test is scored by diff. scorers at same or diff. times

Scanned by CamScanner

28-Nov-2018

(iv) Norms

* Norms (in Ethics) :- specific behaviour for every member of the group & deviation is discouraged.

* Norms refer to avg. performance of the representative sample on a given test. Norms are essential for meaningful interpretation of test scores.

(v) Practicability

- A test should be a practical one from the standpoint of the length of the test, degree of difficulty of the test items, ~~the ease with which it can be~~ or from the standpoint of scoring system, method of administration, etc.

Types of Tests

(i) According to mode of administration

1) Speed Test
2) Power Test
3) Group Test
4) Individual Test

1) Speed Test

↳ clerical speed & accuration Test or DAT (Differential Aptitude Test)

Features

(i) Presence of the time factor
(ii) Items with same degree of difficulty
(iii) Scores are decided by finding the no. of items answered correctly within the given time.

2) Power Test

↳ (i) Items are arranged according to increasing degree of difficulty.
(ii) Time factor is not included (a typical power test)
(iii) Measurement of the underlying ability

Scanned by CamScanner

(iv) In contemporary times, time factor has been included for administrative convenience.

→ We ensure that time factor does not constraint the performance (sufficient amount of time is given)

3) Group Test

↳ administered to more than one individual at a time

eg. Army Alpha & Army Beta

Advantages

(i) economical
(ii) Time effective
(iii) Labour effective
(iv) Norms are easily available

Limitations

(i) spatial problems of individuals many times ignored
(ii) Intensive examination of individual not possible

4) Individual Test

↳ administered to one person at a time

eg. SBIT (Stanford Binet Intelligence Test)

- Advantages
  (i) Comprehensive investigation of an individual
- Limitations
  But (i) it is Time, cost & Labour Intensive (TCLI)
  (ii) Norms availabity is difficult

(ii) According to types of items included

(1) Essay Tests
(2) Objective Tests

- Objective Tests

↳ They have one correct answer that the subject has to provide on his own or to select from the given few

→ Selective → MCQ
→ Supply → fill in the Blanks

Scanned by CamScanner

• Essay Tests aka Free Answer Tests

↳ individual has the independence to answer the questions the way he wants & has the freedom to supply information that he desires. So, he is not restrained by fixed response categories.

They measure individual's (i) originality
(ii) organisation
(iii) sentence construction
(iv) Abstraction & generalisation

Q) Compare & contrast objective & essay type test?

Ans= Similarities

(i) Both want to be objective but both end up having subjectivity.

Also, Q) Just putting the test questions in an objective format does not render the test objective. Discuss.

Even in an objective test, there is subjectivity due to:
(i) statement / stem (ambiguous)
(ii) Distractors (weak)
(iii) Agreement on the correct answer

Thus an objective item test, need not be ~~an objective test~~ having objectivity.

Q) Can we say that the scope of subjectivity in an essay test is higher than that in Objective Test.

Q) Test construction ~~decreases~~ decides the objectivity of the objective type test while test scorer decides objectivity of essay test.

* Assignment of numerical scores is the function of Test scorer in an essay test.

* Objectivity is something ~~wit~~ for which both tests strive but each test suffers from subjectivity:
bluffing → sub. Test Guessing → objective Test

Scanned by CamScanner

*In Objective item test, post construction, scoring is not an issue but in essay test, to identify whether the answer caters to the demand of the question has to be decided by the scorer.

(iii) On the basis of theme measured

(1) Achievement Tests

↳ They measure what the person has learnt or acquired as a result of the training he has received.

eg. CS Optional Exam, DAT (Differential Apt. Tests)

(2) Aptitude Tests

↳ predictive instrument → what a person can achieve in future

→ The ability the person can acquire in time to come

eg. DAT (Differential Apt. Tests)

(3) Intelligence Tests

↳ general aptitude tests

(Apt. test is for a specific ability)

(4) Personality Tests

↳ individual's social and emotional adjustments in his environment.

(5) Interest Tests

↳ measure preferences for specific tasks or activities

(6) Value Tests

↳ measure our generalised & dominant interests.

Scanned by CamScanner

29-Nov-2018

(iv) On the basis of the method of construction used:

1) Teacher made test

2) Standardised Test

1) Teacher Made Test

↳ No explicit principles of test construction laid down.

↳ to bring out individual diff. in classroom setting.

2) Standardised Test

↳ which must meet following requirements/conditions

(i) scoring system should be well defined & clearly laid.

(ii) There should be uniformity in the order of presentation of items and procedures of test administration should be clearly defined.

(iii) items must be clearly worded & there should be no ambiguity in their meaning for the test scorer and test takers.

(iv) there should be satisfactory reliability & validity

(v) there should be the availability of satisfactory norms

(vii) Index of fairness of correct answer through the procedure of item analysis must be established.

Q) How we can progress from teacher made test to standardised test?

Scanned by CamScanner

(v) On the basis of the content of the items or the degree of language used

1) Verbal Tests
2) Non-verbal Tests
3) Performance Tests
4) Non-language Tests

→ differs by the degree of lang. ~~used~~ involved in understanding the test instruction & questions and for answering the test questions.

1) Verbal Test eg. SBIT, CSer. Exam

language is used for:
(i) understanding the test ~~quest~~ instructions
(ii) comprehending test items
(iii) Answering " "

a) Lang. is unique human attributes.
b) One of the 3 instruments used for thinking (image & concepts)

* If we do not use language, many cognitive capabilities could not be understood.

* Non verbal test can supplement but cannot supplant verbal test.

→ There is no substitute of language & hence for verbal test.

⇒ Language & thought influence each other.

2) Non-verbal

Raven's Progressive ~~model~~ matrices

↳ for understanding test items lang. is not required
(for instructions, it is used)
(they are also type of performance test but not pure lang.)

3) Performance Test (occasional use of lang., effort to minimise)

↳ motor co-ordination tested
↳ Alexander Pass Along Test
↳ Stromberg's finger dexterity test

Scanned by CamScanner

(4) Non-language

Like perf. test but no use of lang (complete elimination)

eg. Perf test with no lang like Alexander Pass Along test with no use of lang.

Construction of Psychological Test

Steps:

1) Planning of the test
2) Item writing
3) Preliminary Administration (Experimental Tryout)
4) Reliability of the final test
5) validity of the final Test
6) Norms of the final Test
7) Preparation of Test Manual & Printing of the Test

3 Objectives of a psychological Test

1) To establish a functional ~~id~~ relationship with specified variable or variables.
2) To represent the defined Universe of content
3) To measure psychological Traits

1) eg: variables to measure creativity

	Questions
(i) Div. thinking	10
(ii) Autistic " (Dream like thinking)	5
(iii) Con. "	4
(iv) Personality	2
(v) Openness	2

→ The questions should have a relationship (functional) with the trait.

→ Attributes should be objectively defined to establish relationship.

Scanned by CamScanner

2) Universe of content:- Syllabus

↳ Some questions pertaining to each area must be present

Step 1

Planning of the Test

(i) Define the broad & specific objectives of testing/test.

(ii) Probable length of the test (iii) time duration for the test)

(iv) Nature of items to be used → ~~Nature~~

(v) Method of sampling to be adopted

(vi) Degree of difficulty of items

(vii) Arrangement of Preliminary & final trial administrations

Step 2

Item writing

(i) Lowest Common Denominator (LCD) of the test that can be scored)

(Single task that cannot be broken down into smaller unit → Item)

Characteristics of a good item:

(a) moderate degree of difficulty

(b) High discriminatory power

(c) Phrased in a manner that ~~it is its~~ it presents no ambiguity in its meaning to the test taker as well as test scorer.

(d) Items must measure significant aspects of testee's understanding & knowledge.

(e) Items should not be interlocked (not dependent on previous items for its meaning)

(f) must not encourage any kind of response bias eg. guessing, bluffing, etc.

Scanned by CamScanner

* Requirements for effective item writing

Item writing is a creative art. There are no set rules that can guarantee the writing of effective items. Much depends upon originality, experience, ingenuity, etc. of the item writer. However, there are certain characteristics that a good item writer must possess.

(i) Knowledge (& expertise) about the subject matter on which he wishes to write the items.

(ii) Knowledge about the diff. types of items, their strengths and their weaknesses

(iii) Information about the important characteristics of the subject ~~tests~~ for whom he wishes to write the items.

(iv) He must have good command over the language which is to be used for writing the items.

(v) Must avoid writing interlocked items as well as using the stereotyped words & phrases in the test items.

Types of items

(i) essay items
(ii) objective items

essay items

↳ free answer items
↳ open ended items

1) Assignment of numerical score is in hand of test scorer
2) Objectivity → test scorer
3) Qualities like ~~assignm~~ originality easy to determine
4) Scorer to be an expert

Objective items

↳ fixed category items
↳ close ended items

1) ~~In hands of~~ decided by scoring key.
2) objectivity → test constructor
3) These qualities tough to test.
4) Any lay person

Scanned by CamScanner

5) Tough to achieve objectivity

5) Comparatively, easier to achieve objectivity.

Step 3

* Preliminary ~~Examination~~ Administration (Experimental Tryouts)

→ ~~Giving~~ Administering the test to a group of the target population

→ multiple times the test has to be given

→ Based on the feedback, modifications are made in the test.

* Once the items have been written down and have been modified in the light of the suggestions given by experts, the test ~~as~~ is ready for experimental tryout.

* The objective of the tryout is ~~as~~ to identify the weaknesses inadequacies & ambiguities in the ~~item~~ test items if any

(ii) Determine the validity of each item

(iii) Determine the reasonable time limit for the test

(iv) Determine the appropriate length of the test.

(v) Determine the degree of difficulty of each item of the test

(vi) To determine the inter correlation among the items of the test, so that overlapping items can be avoided.

(vii) To determine the discriminatory power of the items of the test.

* Conrad

↳ pre-tryout (I Preliminary Adm.)

↳ Tryout-proper

↳ Final trial Administration

Pre-tryout

↳ I preliminary Adm.

↳ gross ambiguities in items & instructions & administration

Scanned by CamScanner

* Tryout proper

↳ Data for item analysis

↳ technique that allows us to find why a particular item has not functioned effectively and what can be done to make it relevant & effective.
(to suit our purpose)

Item Analysis

↳ it will allow us to finally select the items to be included in the Test

↳ strength of the distractors will be known

↳ discriminatory power of the items will be reveled

↳ degree of difficulty of the items will be identified

Item Analysis can be done by i) experts
ii) subjects of the test
(based on their performance & its relation to the objective)

* Final Trial

↳ To find out those ~~effects~~ defects which could have gone un-noticed in first 2 administrations.

↳ It is a kind of a dress-rehearsal

↳ After this, no change thereafter

↳ It tells us how effective the test would be when it is administered on the subjects for which it was ~~intende~~ intended.

↳ It will tell about & will help in setting length, discrimitory power, difficulty level, etc.

Scanned by CamScanner

30-Nov-2018

Step 4: Reliability of the final Test

→ consistency of the test reflected in the reproducibility of the scores.

→ A behaviour is said to be reliable if repeated measurements produce the same result

→ In the classical theory of test, reliability is defined using 3 assumptions:

eg. Driving Aptitude Test

→ Highlight what the fundamental properties would constitute driving aptitude.

(i) Distance perception
(ii) Reaction time
(iii) Psychomotor co-ordination
(iv) emotional stability
(v) Knowledge of traffic rules

* If all aspects are taken care of, it increases the length of the test which itself introduces subjectivity.

⇒ No test can achieve 100 % reliability

Assumption in reliability:

(i) Each person to be described by the test has some fixed amount of the attribute of interest. This amount is known as person's true score.

(ii) Every observation of an attribute contains some degree of error.

(iii) Observed score represents both the true score & some error.

The usefulness of the test is just at least in part is judged by the extent to which it is reflecting true score. Reliability coefficient provides an estimate of the proportion of the

Scanned by CamScanner

observed score variance i.e. true variance rather than error variance.
(var. bet scores)
(variance due to errors eg. guessing)

* Ways of Reliability Estimation

(a) Test - Retest Method
(b) Split half method
(c) Equivalent form method (Alternate form Method)
(d) Internal consistency method

(Reliability coeff.)

(a) Test - retest method

↳ Test of Mathematics —15 days→ Test

Sample → same sample

Relative ranks are considered

→ Single Test
→ administered on 2 occasions
→ to the same sample
→ scores are obtained
→ correlated scores
→ coeff. of correlation is reliability coeff.
→ Temporal stability (stability coeff.)

Limitation

Same test, thus, next time scores will be better for all as people already know the questions.

(i) It is based on the notion that physical state, mental state & environment of the subjects would be same on both occasions

(ii) practice & memory method

(iii) cumbersome process
↳ absenteeism

Scanned by CamScanner

If the same people come again & again, they are not representing the population (all people are not so motivated in general population)

(iv) If time period is long & subjects are undergoing maturational changes.

But, here, it is relatively easier to compute reliability.

(b) Split Half Method

Test of 100 questions ⟶ split into 2 equal halves

(i) single administration
(ii) single test
(iii) Data is obtained in single testing

equal halves with respect to homogenity & deg. of difficulty.

⇒ Reliability here measures item equivalence

eg. $\begin{matrix} 1+2 \\ 3-1 \\ 3\times2 \end{matrix} \Big| \begin{matrix} 2+2 \\ 3-2 \\ 4\times3 \end{matrix}$

⇒ Here there is a compromise on diversity

→ Only one test is given, test takers have identified which question belongs to which half.

Limitations

(i) No. of items reduced → diversity compromised
(ii) Nature of item reduces on repetition
(iii) Some factors that are found out in test-retest method cannot be identified here (eg. test taker is unwell)

(c) Equivalent form method (Alternate form Method)

↳ Here two tests → two forms of a single test

Test (A) & Test (B)

→ Administered to the subject within 15 day time period

→ subjects will not have repetition of items

→ Reliability here represents equivalence of items & temporal stability.

Scanned by CamScanner

But to have 2 forms of a test as equal is a big challenge

* It will reduce memory effect but will not eliminate it.

(d) Internal consistency

↳ inter item correlation is a measure of internal consistency → if I get one item correct, what are the chances of getting another item correct

↓

When all items have same homogenity & degree of difficulty (measure same behaviour)

High inter item consistency	lower	
1+2	1+2	1+1
2+2	2×4	12345+6789
3+2	3÷1	
4+3		

⇒ All tests have some correlation between questions eg. quick grasping skills will be ~~same~~ useful in diff. subjects.

⇒ Some consistency will be present even if the items are diverse.

⇒ But we have to avoid very high overlap as we have to test diff. proficiencies → Test will be unidimensional (it will measure single behaviour function)

Factors influencing Reliability

(i) Intrinsic factors

(within the test)

1) Length :- As the length of the test, reliability increases as long it is not very long

2) Homogenity :- High

3) Degree of difficulty :- Moderate

Scanned by CamScanner

4) Scorer's reliability or Rater's reliability → scoring ~~sh~~ system should be well-defined & uniform

5) Discriminatory power of the items :- High but not very high

* Extrinsic factors
(outside the test)

(1) guessing
(outside test as tendency of testee)

(2) Environmental variations or fluctuations

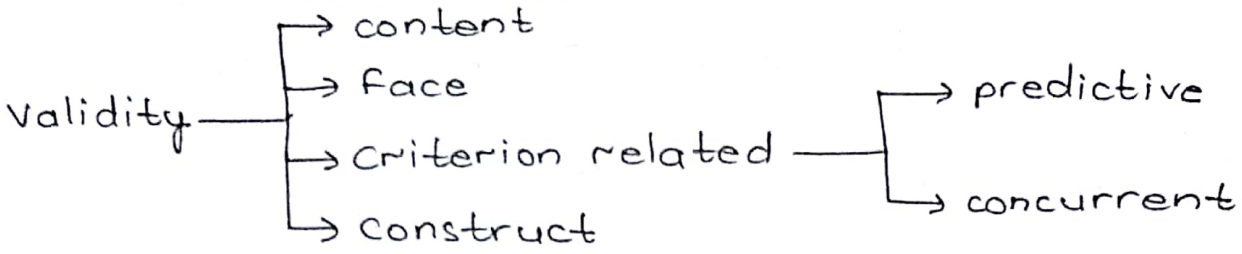
(3) Fluctuations within the examinee
(mental makeup of examinee)

(4) Range of scores or sample heterogenity
(variability in the sample) → High for better reliability

Scanned by CamScanner

01-Dec-2018

Step 5: Validity of the Test



→ Fidelity
→ Purity
⇒ Truthfulness

* degree to which test items measure what they intend to measure

⇒ reliability is the self correlation of the test, validity is obtained by correlating the test scores with scores obtained on the criterion.

⇒ criterion is the best measure of the attribute that the test intends to measure.

Anastasi

~~Fact~~ Validity of the test concerns what the test measures & how well it does so.

(i) Presence of an external independent criterion.
(ii) correlation 'r' between test scores & criterion score

⇒ Validity coefficient

* Characteristics of validity-

(i) validity is not a fixed property of the test because validation is an ongoing process (continuous & regular process)

(ii) validity is not all or none property of the test, rather it is a matter of degrees

Scanned by CamScanner

(iii) Validity is a relative term

(A test is not generally valid. It is valid for a particular purpose)

Types of validity-

(i) Content Validity-

↳ The adequacy with which the test items sample the content area being measured.

↳ Degree to which test items are the representative sample of performance or knowledge being measured.

↳ Achievement Test → very high content validity

↳ Personality Test → very less content validity

Aptitude Test will also have low content validity

⇒ Degree to which intelligence test becomes an achievement test is the success of coaching and failure of the test's purpose.

Q) Test constructor puts all effort to maintain validity & reliability, in diff. methods of testing, it will be diff.?

Ans= It is far easier to obtain high reliability & validity in achievement test than aptitude test as they are dependent even upon the testee.

Aptitude test is future centric, it will have more variables → diff. to obtain reliability & validity.

So, R & V values can be standardised for diff. types of tests.

Q) Can a test with low R & V exist without compromising its utility?

Ans= Yes (something is better than nothing)

eg. Projection Test

↳ measures unconscious thinking

Scanned by CamScanner

→ greater indirectness in the measurement of attribute → lower the subjectivity

⇒ Psychological measurements are always indirect but some measurements are more indirect.
(less indirect are often called as direct)

Two ways to establish content validity –

(i) Expert judgement
(ii) Statistical analysis

Requirements

(i) Before item writing starts, content areas must be clearly specified.
(content area must be well defined so that justice can be done to the various areas included in the syllabus)

(ii) The relevance of the (content of) items should be judged according to response given by the examinees not in the light of apparent relevance of the content themselves.

(ii) Face Validity –

→ degree to which test items measure what they appear to measure.

→ degree to which items appear to measure the variable being measured.

→ If items ~~the~~ appear to measure, they might measure or not → but they should ~~be~~ appear to measure, subjects co-operation can be obtained.

* It is not validity in the technical sense, it is more about superficiality rather than actuality.

Scanned by CamScanner

(iii) Criterion Related Validity

The term Criterion related validity usually refers to the extent to which a measure of an attribute demonstrates an association with some independent or external indicator of the same attribute. This external indicator called as criterion often represents the behaviour we are actually interested in. Criterion related validity is obtained by comparing the test scores with the scores obtained in the criterion available at present or ~~exist~~ to be available in future.

→ Predictive Validity of the test is the extent to which it is efficient in forecasting & differentiating behaviour or performance in a specified area under actual living & working conditions. In predictive validity the test is correlated against the criterion that will be available sometime in future.

→ Concurrent validity: In this, the relation bet. the test & criterion scores is established at the same time.

Q) When criterion is req. in any validity, then why 'criterion' term associated with one type of validity?

Ans= Criterion is a ~~feat~~ feature of content validity but there it is a specific requirement (content is criterion there).

⇒ personality & Aptitude tests :- performance is not judged on the basis of knowledge (criterion Rel. validity)

Scanned by CamScanner

eg. Local Test & standard Test, comparision of scores obtained by individualsin both the tests.

* predictive :- Taking (time gap) test ~~(time gap)~~ of people who later prove succesful.
* concurrent :- Take tests of people already doing that work (No time gap)

(iv)
* Construct Validity-

Scientifically speaking, construct validity is one of the significant advances of modern measurement, theory & practice. It is a signi advance because it unites psychometric notions with theoretic notions. The measurement expert when he enquires into the construct validity of the test wants to know what psychological & other properties explain the variance of the test. He wishes to know the meaning of the test. He wants to know the factors that underlie test performance. It is for this reason, it is also called as factorial or trait validity. Construct validation is a more complex & difficult process than content & criterion related validation, therefore, an investigator decides to compute construct validity only when he is fully satisfied that neither any valid or reliable criteria is availdde to him and nor the universe of content is entirely satisfactory to define the quality of the test.

→ Anastasi has defined construct validity as the degree to which test may be said to measure theoretical construct or trait. Thus construct validity of a testing instrument is based on the determination of the degree to which the test items capture hypothetical quality or trait.

Scanned by CamScanner

A construct is a sort of a concept which is formally proposed with a definition & is related to empirical data. A construct indicates a hypothesis which will tell us about the behaviours that will correlate with one another in the studies of individual differences.

3-Dec-2018

Steps in construct validity-

Step 1:- Defining the construct

Step 2:- Specifying all the possible measures of the construct

Step 3:- Determining the extent of correlation bet. the measures of the construct

Step 4:- Determining whether or not these measures successfully measure the construct.

Acc. to Campbell & Fiske, successful construct validation requires convergent & discriminant validation. For successful conv. validation, it is essential that the test correlates with all those measures with which it should correlate and when the test successfully does so. In other words, correlates with its expected referents, the test is said to have conv. validity. Likewise, it is also necessary, the test must not correlate with the measures with which it should not correlate & when the test correlates poorly with the measures which it should not correlate, the test is said to have discriminant validity.

* Construct validity → when measures for content and criterion validity are not present.

** Any test which must have come should have necessarily undergo construct validity.

Scanned by CamScanner

* Factors influencing Validity

(i) Length of the test (moderate)

(ii) Difficulty level of the items → moderate diff.
→ no abrupt change in the degree of difficulty.

(iii) Discrimitory power (higher)

(iv) Ambiguous directions are likely to lower the validity.

(v) Sample heterogenity (will increase the validity)

(vi) Socio-cultural differences
↳ a psychological test that is designed to measure individual differences of one cultural group when is used for a different cultural group is likely to produce errors and its validity is likely to get compromised.

* Relationship bet. Reliability and Validity:

(i) R & V are two sides of the same coin — test efficiency

(ii) Reliability is the self correlation of the test & validity is the correlation with external independent criteria.

(iii) A test with poor self correlation is not expected to yield high correlation with external independent criteria.

~~Qs / Es~~

* Is reliability a pre-condition for validity ?

Ans = ~~Yes~~ Not necessarily, reliability is a pre-condition for validity but if a test has high validity, it will have high reliability if the test is homogenous test.

In case of heterogenous test, high validity but less reliability eg. CAT having Quant, LR, DI.

Scanned by CamScanner

→ correlation bet. diff. items will be poor in heterogenous test.

* If both V & R involve a trade off, in that case a middle path → without compromising relability achieve maximum possible validity

• In this case, Reliability → internal consistency.

If all items have a) nearly equal deg. of diff
b) homogeneous
} → validity will be compromised

• Thus, a good test constructor needs to aim for a balance between the two.

* High R & V → (i) SBIT (Stanford Binet Intelligence Test)
→ (ii) Wechsler Adult Intelligence Scale (WAIS)

* Low R & V → (i) RAT (Remote Association Test) {creativity test}

→ (ii) Seashore Musical Aptitude Test (MAT)

*(Not necessarily give these examples for low R & V, general examples can be given)

Norms of the Final Test

*Norms refer to average performance of a representative sample on a given test. Norms are essential for meaningful interpretation of the test scores. There are diff. types of norms:

(i) Age equivalent Norms

Eg. ↳ Average performance of a representative sample of a certain age level on the measure of certain trait or ability. They are suited for the abilities which increase systematically with age eg. height, weight, etc.

Scanned by CamScanner

(ii) Grade equivalent Norms

↳ avg. performance of a representative sample of a certain grade or class.

(iii) Percentile Norms

↳ Very popular norm used in psychological & educational test. A percentile norm for each raw score indicates the % age of standardised sample whose score falls below that raw score.

Preparation of the Manual & Reproduction of the Test

In a manual, the test constructor reports properties of test, method of scoring, administration of the test, Time limit & the norms & the references used. He also reports about the instructions to be given and the details of the arrangement of the items and finally depending upon the req. of the test, he orders for printing of the test.

* Uses of a Psychological Test

(i) Personnel Selection & Training

(ii) Clinical Diagonosis & Treatment

(iii) Vocational guidance & counselling

(iv) Psychological tests can be used to collect data for policy formulation & decision making

(v) Psychological test can be used for SWOT Analysis. It can help to promote individual's self understanding by helping him develop insight into his strengths & weaknesses and informing him about how he can utilise them to his advantage.

Scanned by CamScanner

(vi) Psychological test can also be used for the selection of experienced candidates. Some candidates bluff their way through application blanks but bluffing a psychological test is very tough.

* Misuses of a Psychological Test:

(i) Invasion of privacy -
↳ max. in projective Test (without telling the testee)

(ii) Blackmailing by the Psychologist

(iii) Over promotion
↳ of the effectiveness of the test

A general danger in psychological testing is ~~it~~ the possibility of it being oversold~~ing~~ to it prospective clients. Many testing org. exists & they make over enthusiastic claims and these claims are more likely to be made when unprofessional psychologists run these organisations. Psychological testing is a good thing but like all good things, it should not be over done. When incompetent psychologists carry out psychological testing, they fail to maintain high standards req. for Psychological testing. Authorities must ensure that no exaggerated claims are made regarding Psychological tests. A psychological test should be put to use only when its credentials are well proved and its reliability & validity is ascertained. This would imply that only few quality tests are produced but it will ensure that the ones that are produced turn out to be truly ~~objective~~ objective instruments.

(iv) Use of deception in testing
↳ making the persons believe in something that is not true.
(deception by participant is limitation not a misuse)

Scanned by CamScanner

(v) Breach of confidentiality-

↳ Results of Psychological tests should not be made public and if the data need to be released for public, it should be done only after obtaining permission from the subjects tested.

(vi) secret techniques & scoring methods

In the profession of Psychology the use of secret formulae & scoring methods is both unethical & unprofessional. Such people should not be allowed to use Psychological test until the prove their claim.

(vii) Competence of the testers → only competent people should be allowed to conduct tests.

Incompetent testers besides failing to make correct diagonosis are also likely to mislead the subject thereby causing in him unnecessary confusion. Therefore any organisation carrying out Psychological testing must ensure that only qualified people are allowed to take Psychological Test.

Scanned by CamScanner

4-Dec-2018

* Limitations of a Psychological Testing

(i) Culture bias → statistical difference in the validity of the test for diff. cultural groups.

→ most commonly found in Intelligence Test.

(i) Language

(ii) Test constructor → reflected in choice of items

→ Culture fair & culture specific tests

*A test validated for one cultural group and used for a diff. cultural group → that test will be less valid for the diff. cultural group.

(ii) Incompleteness of Psychological Tests :

eg.
There can be disagreement on the factors that influence a behaviour, on the weights of factors, etc.

(iii) Indirectness of measurement

(iv) Lack of absolute zero

(iii) Indirect: Attitude (to be measured) → Behaviour (measured)

[Env. factors to be held constant]
[Internal disturbances to be held constant]

∴ Going from behaviour to attitude will involve some subjectivity.

(iv) how much more or how much less

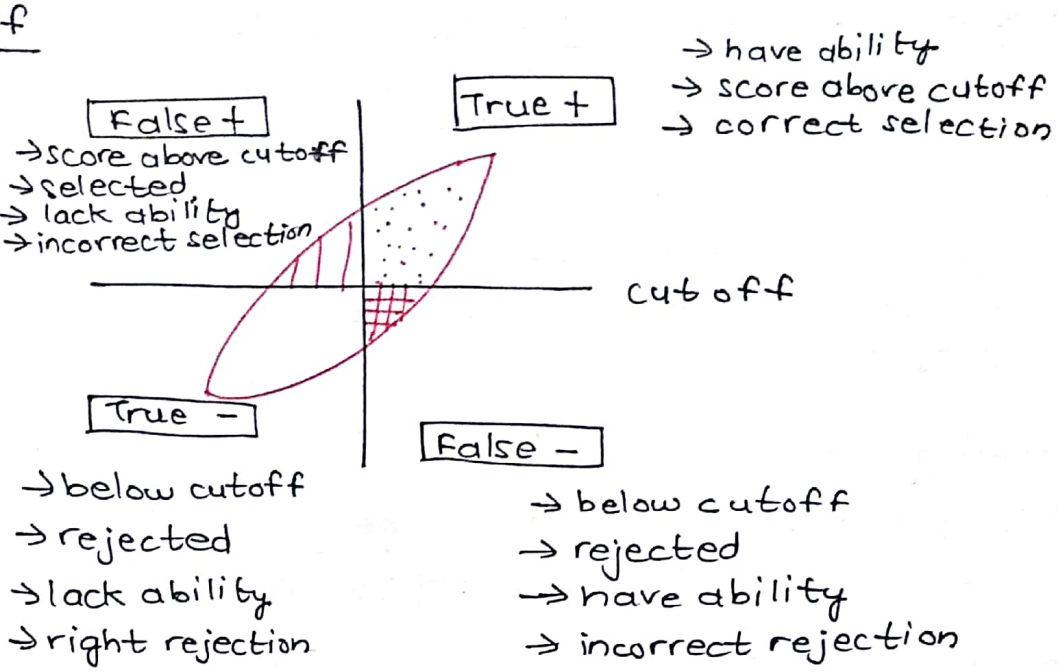
(v) Categorisation

↳ ind. diff. can be better addressed → but if categorisation is ~~indirect~~ incorrect ↓ Individual will be put in unfavourable category ← Social Stigma ← lower motivation & self confidence

Scanned by CamScanner

eg. In jails, people of minor crimes are also kept with hard-core criminals.

(vi) Cutoff



⇒ To ensure only True+ get selected, test will become TCLI, time, cost & labour Intensive.

∴ Thus, placing cutoff depends upon the objective of test. (Whether it strictly wants true +)

(vii) Invasion of privacy-

(viii) Response Bias

↳ the bias of the individual in responding to ~~is~~ test questions → as per his expectations & not ~~a~~ what the test wants to measure

eg. if ~~on~~ the individual thinks that an honest response would portray him in -ve light.

* Response bias is one of the several factors which invalidates the test scores. When a response to an item tends to be altered by an examinee in a way that it indicates something other than what the test intended to measure, it is called as the response bias.

Scanned by CamScanner

Response set is a kind of response bias which occurs as a result of the assumption of mental sets by the examinee. Acc to Cronbach, response bias is a tendency of a person to give diff. responses to the items than he would when the same is presented in diff. form.

*Types of response bias:

(i) Deviation set

It refers to the tendency to give very unusual or uncommon responses to the items in a test. (in projective test)

(ii) Social Desirability & Undesirability Tendency-

↳ (faking good / faking bad)

* faking good or social desirability tendency is a feature of personality test and service selection tests. It refers to the tendency on the part of the examinees to give answers in a way that create favourable impression.

* faking bad on the other hand is the tendency to give answers that may create unfavourable impressions of self. Such tendency is also called as social undesirability tendency

(iii) Acquiescence set

↳ It refers to the tendency to respond in the same systematic manner to all the items of the test. It operates in personality & interest inventories eg. subjects might give 'yes' responses to all the items of the personality test.

↳ for 'No' response → Negativistic

Scanned by CamScanner

(iv) Evasiveness

It is the tendency of the subject to avoid two fixed response categories & answer in favour of doubtful category.

(v) Tendency to work speedily

The tendency to work for speed rather than accuracy is a common response set in a test designed to measure ability. Subjects having ego oriented motivation go through the test paper rapidly without caring much for accuracy. This tendency lowers the validity of the test as it fails to reveal the true picture of the traits being measured.

* Consequences of Response Sets

(i) They contribute to error in measurement as they introduce in the test the factors that happen to dilute the objectivity of the test.

(ii) They reduce the range of individual differences.

*How can we remove response biases:

(i) By modification of the test instructions.

(ii) Minnesota multiphasic personality Inventory (MMPI) use of correction keys

(iii) By redesigning the test items

eg. instead of Yes/No items use MCQs (2 point scale)

(iv) By adapting the degree of difficulty of the test.

(v) The use of more appropriate scoring techniques & formulae.

Scanned by CamScanner

Ethical Issues in Psychological Testing

* To prevent the possible misuse of psychological test APA (American Psychological Association) has adopted an ethical code which devotes considerable space to test distribution & its use.

Some aspects of this code are:

(i) The sale & distribution of the tests should be restricted to the qualified users. Who will be qualified will however vary acc. to the type of the test & the background requirement of the psychologist.

(ii) Test scores should be released to those individuals who are qualified to interpret them.

(iii) A person should not be subjected to testing under fall pretences

eg. A psychologist in the industry may be placed in a conflict situation bet. his loyalty to the person being tested and his loyalty to his employer. In many cases, the tester has to guarantee anonymity to the person taking the test to secure his co-operation only to find the pressure being put on him to release the scores of certain individuals who have been tested. It is the duty as well as the responsibility of the psychologist to keep the testee informed about the purpose of testing and the uses to which his test scores will be put into.

(iv) Test should not be released for public use without objective data to support its merit. If the test has been given for research, this should clearly ~~be~~ be indicated to all concerned

(v) Test manuals containing the data relating to reliability, validity and other psychometric properties should be made available to all concerned &

Scanned by CamScanner

this information should be available for all the tests and should be continuously revised.

(vi) Test or a part of it should never be published or presented in public media such as magazines, newspapers, etc.

(vii) All testing should be carried out under standardised conditions & under the watchful eyes of qualified examiner.

(viii) One should never prepare or coach a person for a particular exam as doing this invalidates the test score as it provides an extra edge to all those who are receiving coaching in comparision to those who are not fortunate enough to receive it.

Scanned by CamScanner